REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE 02/26/03	3. REPORT TYPE AND DATES COVERED Final Progress Report Phase I 07/01/02 -- 02/26/03

4. TITLE AND SUBTITLE
Molecular Signature of Biological Pathogens

5. FUNDING NUMBERS
Contract # DAAD19-02-C-0054

6. AUTHOR(S)
Guck T. Ooi, Ph.D.; Sun H. Paik, Ph.D.; Earl W. Ferguson, M.D., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Sun BioMedical Technologies, Inc.
1539 N. China Lake Blvd. , PMB231
Ridgecrest, CA 93555

8. PERFORMING ORGANIZATION
REPORT NUMBER 6 (FINAL PROGRESS

20030317 125

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

43820.1-LS-BI

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Peripheral blood mononuclear cells (PBMCs) were infected *in vitro* to develop a model for studying infections of high interest with regards to bioterrorism threats. PBMCs were infected with *E.coli, B.subtilis* and *B.cereus*, and gene expression analyzed by DNA micro-array technology. Each group had clearly different genetic expression profiles. Random Forest Predictor classified Control, *E.coli, B.subtilis* and *B.cereus* groups with only one misclassification in 34 samples (test set accuracy=97%). A list of the 20 most important predictor genes was developed using stepwise linear discriminant analysis. *In vivo* responses to initial Anthrax vaccinations and *E.coli* urinary tract infections were compared to *in vitro* responses to *B.cereus* and to *E.coli*, respectively. *E.coli* urinary tract infections differed from controls, but did not group with *in vitro* infections, suggesting a limited systemic response. Anthrax vaccinations produced a distinct response, different from controls but similar to the *in vitro* responses from *B.cereus* infections. These findings suggest that this model will predict systemic infection responses to bioterrorism pathogens of interest. Analysis of proteomic responses to identify key markers for specific infections provided critical information to guide *in vivo* analyses that will be expanded to a larger group of subjects with systemic infections in Phase II.

14. SUBJECT TERMS
microbial pathogens, immune response, genomics, molecular markers, DNA microarray, bioterrorism, Bacillus

15. NUMBER OF PAGES
19

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION ON THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

REPORT DOCUMENTATION PAGE (SF298)
(Continuation Sheet)

MOLECULAR SIGNATURES OF BIOLOGICAL PATHOGENS
Phase I Final Report:

1) Foreword:

DoD CBD 02-100 Objectives

The main objectives of the DoD CBD 02-100 project are to establish and identify the specific molecular signatures of different pathogens, and to determine whether these signatures can be used to forecast/predict expected early molecular markers of *in vivo* infection with biological warfare agents of high interest with regards to bioterrorism threats (Centers for Disease Control and Prevention [CDC] Category A biological agents).

Our Research Work

While the above CBD objectives focuses on *in vivo* studies to determine the response of normal volunteers to chance infection by specific bacterial or viral pathogens to be identified after infection occurs, we felt that baseline *in vitro* basic studies should be accomplished first, together with some complementary *in vivo* studies to identify key issues associated with *in vivo* work. This combined *in vitro/in vivo* has the following advantages:

1. Rapid optimization of critical experimental parameters involved in acute infections (such as, time course of specific infections) and characterization of specific molecular responses and early molecular markers that are expected *in vivo*

2. Characterization of molecular responses to infection and early molecular markers for pathogens that are not expected to occur and can not be tested in normal populations, but are of high interest with regards to bioterrorism threats (Centers for Disease Control and Prevention [CDC] Category A biological agents: *Bacillus anthracis*, *Clostridium botulinum* [botulism], *Yersinia pestis* [plague], *Francisella tularensis* [tularemia], pox viruses, and hemorrhagic fever viruses);

3. Prediction of early molecular markers that would be generated by *in vivo* responses of healthy, human subjects to biological warfare agent exposure;

4. More cost-effective, focused application of expensive DNA microarray technologies in development of the envisioned database of the human genomic response to various pathogens;

5. More focused and simplified *in vivo* studies on human volunteers.

Objectives of Our Phase I/II SBIR Research Work

The specific objectives of our Phase I/II research work are consistent with the DOD CBD 02-100 objectives, and includes the following:

1. Identify and characterize genetic responses to pathogen exposure at a genomic level.
2. Identify early molecular markers of biological agent exposure.
3. Develop a database of human responses to various pathogens so that exposure can be determined and the agent can be accurately identified within minutes or hours of infection.
4. Determine the host gene expression "signature" of microbial pathogen exposure and identify distinct host responses to different pathogens.
5. Train a Random Forest Predictor [RFP] algorithm (and or other algorithms, such as Support Vector Machine [SVM]) to allow accurate identification of an unknown pathogen exposure

2) Table of Contents:

3) List of Figures: (Figures attached as Appendix)

4) Statement of Problem Studied:

Infection by a microbial pathogen triggers a complex and distinct set of coordinated cellular and systemic events that result in the host-defense response. Interactions between a host and microbial pathogens are diverse and regulated in specific patterns by unique molecules and mechanisms involving activation of transcriptional events of innate and adaptive immunity [1]. Individual pathogens develop their own strategy for survival in host target cells and may elicit a specific host response besides the broad and generic local recruitment of leukocytes or T lymphocyte subsets and secretion of cytokines that promote cellular and humoral immunity.

The complex interaction between microbial pathogen and host in infectious disease processes can be explored by analysis of gene expression to provide details of the early molecular events that follow infection and to better understand their regulation [2,3]. The knowledge of human genomic sequences is just the starting point for unraveling the complexities of this host-pathogen interaction. Infection of a host by pathogenic bacteria involves changes in the physiology of both host cells and invading microbial pathogens. These physiological changes are due to gene expression changes that reflect and characterize an ongoing infectious process and are unique to specific pathogens. The host profiling of gene expression by DNA microarray hybridization may identify gene expression signatures unique for each pathogen and may identify functions of genes not previously implicated in the response to infection. Patterns of host gene expression response to different pathogens have been described for many virus and bacteria but have been limited to few well-known cytokines that are strongly induced in response to different inflammatory stimuli [4]. High-density DNA microarrays can identify genome-wide transcriptional events that underlie host response to microbial pathogens.

Profiling gene expression patterns of host cells before and after specific infections will provide better understanding of differential microbial pathogenesis and may provide novel tools for early diagnosis and clinical management of specific infectious diseases, including the identification of new therapeutic targets. Traditional diagnostic approaches require isolation of the etiologic agent or measurement of antibody response to a specific pathogen. In this project we propose to create a host gene expression "signature" to early microbial pathogen exposure and identify distinct molecular level host responses to different pathogens that do not require isolation of the pathogen or waiting for the host antibody response.

Microarray technology can quantify the differential expression of thousands of genes in various pathogenic states. Distinct host gene expression "signatures" can be used as diagnostic markers of infection for early detection of exposure to pathogens and to determine time of exposure.

5) Summary of Most Important Findings:

Phase I research was restricted to showing feasibility of analyzing the early differential immune response of PBMCs to *Bacillus cereus, Bacillus subtilis,* and *Escherichia coli* and to validate *in vitro* data by detecting a differential immune response to *Bacillus anthracis* vaccinations and *Escherichia coli* urinary tract infections by analysis of blood samples. Investigation of other pathogens to generate a more comprehensive database of human response to various types of Gram-positive and Gram-negative bacteria and viruses in a larger group of subjects with multiple sampling periods will be undertaken in Phase II.

a) Processing of Samples: The proof-of-concept experiments were carried out *in vitro* for closer control of infection conditions and time post-infection. To demonstrate that the *in vitro* infection reflect or closely mimic the *in vivo* infection, we analyzed and compared gene expression profiles of PBMC from patients with urinary tract infections (culture proven to be *E. coli*) and PBMC infected *in vitro* with *E. coli*. This approach will allow us to test the host response to many virulent pathogens (including biowarfare microorganisms) to obtain a "fingerprint" for specific infectious agents. In parallel, experiments were done also with an opportunistic pathogen *Bacillus cereus* (genetically related to *B. anthracis* with 92.2 – 99.6% DNA sequence identity and 96.5% amino acid sequence identity) and ubiquitous soil bacterium *Bacillus subtilis* 168 (evolutionary divergent *Bacillus* strain). These two strains were chosen to demonstrate differential host discrimination between related bacterial species (*B. cereus* vs. *B.subtilis* 168) and *E. coli* was chosen to demonstrate host discrimination between genetically and evolutionary unrelated species (Gram-positives and spore-forming *B. cereus* and *B. subtilis* 168 vs. Gram-negative *E. coli*).

Blood samples were collected from healthy, genetically diverse anonymous volunteers similar to the population found in the U.S. Armed Forces. Based on control experiments, blood sample volumes (120 ml) were increased and the number of blood donors decreased. All *in vitro* studies (*B. cereus*, *B. subtilis 168*, *E. coli* and Control*)* were completed on each sample to decrease the likelihood of individual variability between groups. Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll-Paque and cultured with *Bacillus cereus* or *Bacillus subtilis,* or *Escherichia coli*, for comparison to control cultures. To minimize the initial costs related to DNA microarrays, only a single concentration of bacteria load for each group of PBMC infections was tested (10:1 or lower [1:1 for *B. cereus*] multiplicity of infection for 3 h in CO_2 incubator at 37 ^{0}C). The MOI for *B. cereus* was decreased because fast growth and attachment to PBMCs resulted in cell lysis and poor quality of RNA at higher MOIs. After incubation, cells were harvested, washed and processed for total RNA extraction using the RNeasy Total RNA Isolation kit (Qiagen) recommended by the Affymetrix protocol. The quality of the RNA samples was documented by agarose gel, absorbance ratio at 260nm/280nm, and Agilent RNA Analyzer. All RNA samples submitted for DNA microarray analyses passed stringent quality controls.

b) Initial DNA microarray analyses: Working DNA microarray data sets comprised of 42 samples divided into 6 groups as summarized in Table 1.

Table 1. Microarray data sets used for analysis to determine whether gene expression profiling can be used to identify pathogen types.

Treatment Name	Number of Samples	Comments
Ctrl	12	Non-infected control group
E. coli	7	*In vitro* samples infected with *E. coli*
B. cereus	7	*In vitro* samples infected with *B. cereus*
B. subtilis	6	*In vitro* samples infected with *B. subtilis*
UTI	2	*In vivo* samples from patients with Urinary Tract Infection, confirmed to be due to *E. coli*
AV	4	*In vivo* samples from volunteers 24 h after Anthrax Vaccination
UnkA, UnkB, UnkC, UnkD	4	Masked *in vitro* and *in vivo* samples included to test the precision of gene expression profiling in identifying infection type

To determine treatment effect (pathogen type) on global gene expression profile, unsupervised learning analysis of the data was performed. Hierarchical clustering analysis using all 22,215 genes showed that samples cluster into 6 groups determined by pathogen type (Control, *E. coli, B. cereus*, *B. subtilis,* UTI and AV). Similar conclusions were obtained when data was analyzed using the 5000 and 1000 most varying genes based on the coefficient of variation (Figure 1, see Appendix). Multidimensional scaling plots confirmed inferences made from hierarchical clustering analysis (Figure 2, see Appendix). Results confirmed that changes in gene expression profiles are different for different pathogen types, and can be used as signatures for identifying pathogen exposure. There was no global gender, age, or race effect using unsupervised learning analysis (clustering and multidimensional scaling plots), although this may be due to small sample size. This issue will be further addressed in phase II with larger sample sizes.

Several 2-group comparisons using the t-test filtered out genes that were significantly different at p-values equal to or smaller than 0.01. In each file, the genes were sorted by the t-test statistics, the larger the absolute values of the t-statistics, the more significant the genes. Shorter gene lists were available from the sorted list by setting more stringent criterion, i.e., p=0.005, p=0.001, etc.

Differences between *in vitro* infection groups: As stated above, pathogen type was clearly separated by global gene expression profile. Using t-test for 2-group comparisons (infected group vs. control), there were significant differences between each infected group compared to controls. At the P<0.01 level, a series of gene list were compiled for different groups. The following number of genes were different (P<0.01) for each comparison:

4043 genes between Ctrl vs. all *in vitro* infected groups (*B. cereus, B. subtilis* & *E. coli* combined)
2958 genes between Ctrl vs. Bacillus groups (*B. cereus* & *B. subtilis* combined)
2464 genes between Ctrl vs. UTI

Differences between Gram-negative bacteria (*E. coli*) vs. Gram-positive bacteria (*B. cereus* & *B. subtilis* combined): Unsupervised learning analysis indicated that there were significant differences between Gram-positive and Gram-negative bacteria. These two groups clustered separately, and t-test comparison at p<0.01 level filter out 1339 differentially expressed genes.

Differences between *in vitro* vs. *in vivo* infected groups: Comparisons were made between *E. coli* vs. UTI and *B. cereus* vs. AV to determine whether *in vitro* infection reflects similar or related *in vivo* infections. Blood samples were collected from women with UTIs and processed for DNA microarray analyses as described above. The samples from women with culture proven *E. coli* UTIs were sent for analysis with *E. coli in vitro* samples. Half of the UTI samples initially sent for processing were lost in sample processing at the DNA Microarray Facility at UCLA. The two remaining UTI samples did not group with the *in vitro E. coli* samples, but did separate from controls using the unsupervised learning analysis. Based on those results, additional UTI samples were not processed since UTIs appeared to act as "localized infections" rather than systemic infections and did not appear to generate sufficient systemic changes to completely mimic *in vitro* responses. Nevertheless, UTI group can be clearly differentiated from Ctrl group. Two group comparison using t-test at p<0.01 level indicated 2464 genes that are differentially expressed during UTI. The Correlation Matrix of UTI samples to *in vitro E. coli* samples showed overall correlation of 0.86 (= 74 % Similarity).

For AV samples, blood samples were collected 24-26 hours after initial anthrax vaccinations in 5 subjects and processed for DNA microarray analyses. Five samples were sent for analysis as post-anthrax vaccination samples (out of this five, one sample was masked as UnkB). In an unsupervised learning analysis, all 4 AV identified samples clustered together but away from *B. cereus* samples indicating that there are differences between *in vivo* response to *Anthrax* vaccinations and *in vitro B. cereus* infected samples. This is validated in a t-test comparison, where 2819 genes were obtained that showed highly significant (p<0.001) changes. Nevertheless, all AV samples can be clearly differentiated from Ctrl group. Using p<0.001 cut-off level, we cataloged 1822 genes that are differentially regulated due to anthrax vaccination. This difference is somewhat expected as anthrax vaccination (soluble protein fraction) should not be expected to elicit exactly the same immune response as a live *B. anthracis* infection and *B. cereus* is not identical to, but similar to *B. anthracis*. Even so, the Correlation Matrix of *Anthrax* vaccination samples to *in vitro B. cereus* samples showed overall correlation of 0.89 (= 80 % Similarity).

c) Molecular signatures for specific infection groups: Clustering analysis and multidimensional plots together with pair-wise t-test comparisons identified a list of genes whose expressions were significantly altered in each pathogen groups. Using these gene lists, a supervised analysis prediction method (Random Forest Prediction method developed by L. Breiman [5]) was used to determine the pathogen status of known 36 samples plus four unknown samples (Control, *E.coli, B.subtilis,* and *B.cereus*). When the Random Forest Parameter entry was set at the 2000 most important genes, the predictor was 97 % accurate for classifying *in vitro* samples and 92% accurate for combined *in vitro* and *in vivo* AV samples.

Table 2. Classification tables by Random Forest Prediction:

Treatment Group	Treatment Name	Sample Number	Mis-Classification	Correct Classification	% Correct
1	Control	12	1	11	91.70%
2	*E.coli*	7	0	7	100%
3	*B.subtilis*	6	0	6	100%
4	*B.cereus*	7	0	7	100%
	TOTAL	**32**	**1**	**31**	**96.90%**

Treatment Group	Treatment Name	Sample Number	Mis-Classification	Correct Classification	% Correct
1	Control	12	2	10	91.70%
2	*E.coli*	7	0	7	100%
3	*B.subtilis*	6	0	6	100%
4	*B.cereus*	7	0	7	100%
5	AV	4	1	3	75%
	TOTAL	**36**	**3**	**33**	**91.70%**

The Random Forest Predictor calculates not only measures of gene importance, but also the most important genes for predicting infection status. From the list of the 200 most important genes, a final list of the 20 most important genes was determined using stepwise linear discriminant analysis. The 20 most important genes lead to a perfect separation of the different infection groups (Figure 3A, 3B, 3C, see Appendix).

d) Random Forest Predictor for determining pathogen status of masked samples: Besides clustering accuracy as a measure of determining the precision of the Random Forest Predictor, 4 unkown samples were included in the microarray analysis, that remained unkown to both the microarray technician and the statistician performing the data analysis. The Random Forest Predictor was able to identify accurately UnkA, UnkB, UnkC, and UnkD to be B. subtilis, AV, E. coli, and B.cereus respectively – 100 % accuracy. This "blind" testing confirmed that changes in global gene expression profiles can be used accurately to identify exposure to biological pathogens. The classification probabilities of 4 masked samples are shown in Table 3.

Table 3. Classification Probabilities:

Sample ID	AV	*B. cereus*	*B. subtilis*	Control	*E. coli*
UnkA	0.1816	0.1370	**0.3952**	0.2098	0.0764
UnkB	**0.5968**	0.0258	0.0188	0.3316	0.027
UnkC	0.0888	0.0944	0.1452	0.1362	**0.5354**
UnkD	0.1308	**0.3966**	0.2496	0.1096	0.1134

e) Proteomic analysis of samples: Preliminary proteomic analyses identified specific qualitative and quantitative protein changes when PBMC cultures were stimulated *in vitro* with *E. coli, B. cereus* or *B. subtilis* bacterial strains. Non-infected and infected PBMC culture supernatants were analyzed to determine differentially secreted cytokines and/or lymphokines that can be detected by HPLC and two-dimensional (2-D) gel electrophoresis. This effort was directed at proteins secreted in plasma to identify protein markers that could be used for rapid detection by biosensor technology. Clear differences in HPLC profiles could be seen among non-infected control samples and samples from *E.coli, B. cereus* and *B.subtilis* infected culture supernatants. Proteins were separated from culture supernatants by analytical reverse-phase-HPLC with a Vydac C18 column and three-step linear gradients. Although some differences were observed among subject samples, there were characteristic protein patterns differences between control samples and samples from specific infections (Figure 4A, see Appendix). For example, all six PBMC cultures infected with *E. coli* showed a peak eluted at 17 min (absent in control samples and *Bacillus* sp. infected samples) and an inverted double peak eluted at 8 min of reverse-phase HPLC. Culture supernatants of PBMC infected with *B. cereus* and *B. subtilis* also showed differential protein secretion patterns compared to controls and *E. coli* infected samples. Figure 4B (see Appendix) shows distinct HPLC profiles of serum samples before and 24 h after *Anthrax* vaccination in the same subject.

For better separation of secreted proteins, non-infected and infected PBMC culture supernatants were analyzed by 2-D gel electrophoresis. Culture supernatants were concentrated 5 to 10 times using a 3K Dalton cut-off protein concentration device (NanoSep) to improve the visualization of low abundance proteins. To improve the fractionation of serum proteins present in samples, albumin was removed using SwellGel (Pierce) resin columns. Although it improved the separation of protein spots in the second dimension, the SwellGel blue resin also trapped other serum proteins. Loss of bands by 1-D SDS-PAGE electrophoresis and protein spots by 2-D electrophoresis gels was observed when samples were compared before and after albumin removal. 2-D electrophoresis was performed using Bio-Rad System and reagents. The best sensitivity was obtained using fluorescent Sypro Ruby stain (rather than Silver Stain Plus) and improved Bio-Safe Coomassie Blue (Bio-Rad). Unique proteins (spots) were identified by gel comparison using Quantity One Analysis software (Figure 5, see Appendix). The analysis of HPLC profiles and 2-D electrophoresis gels demonstrate that PBMC cultures express and differentially secrete protein markers in response to specific infection. In Phase II, these unique protein spots will be further identified and characterized with 2-D image analysis PDQuest software. Downstream protein spot identification after excision from gels will be obtained by peptide mass fingerprint analysis using ESI-MS-MS mass spectrometry.

Western blots were used to evaluate correlation between gene expression and protein levels. Based on gene expression data, three cytokines with commercially available antibodies were tested. Good correlation was demonstrated between gene expression levels and protein levels of TNF-α and IL-4 (Figure 6, see Appendix). However, no correlation was found with cytokine Amphiregulin, despite relatively high gene expression levels in *B.cereus* in comparison to *B. subtilis*, *E. coli* and Control. Trace amounts of Amphiregulin were detected in two of 6 cultures with *E. coli* (Figure 6). Amphiregulin was not detected in any of 6 samples each of control, *B. cereus*, and *B. subtilis* groups. According to gene expression data, Amphiregulin levels comparable to IL-4 levels shown in *E. coli* group should have been detected in *B. cereus* culture supernatants by Western blot. These studies demonstrate that gene expression data can guide the study of responses to specific infection but complementary proteomics data is necessary for identification of unique sets of protein markers of specific infections.

f) Conclusions:

Phase I demonstrated that unique differential genetic expression profiles can be identified and characterized for specific pathogen exposures and that distinct molecular markers of infection can be identified within 3 hours after *in vitro* exposure and 24 hour after *in vivo* exposure (Anthrax vaccination). This demonstrates the feasibility of establishing a combined *in vitro/in vivo* database of differentially regulated genes for each pathogen type to identify distinct host responses to different pathogens. This database will assist in prediction of responses to biological agent exposures that cannot be tested *in vivo* and are not usually encountered in human subjects (such as, CDC Category A biological agents: *Bacillus anthracis, Clostridium botulinum* [botulism], *Yersinia pestis* [plague], *Francisella tularensis* [tularemia], pox viruses, and hemorrhagic fever viruses). Training data sets for accurately identifying human responses to various pathogens were used with the Random Forest Predictor to accurately identify unknown samples (*E.coli, B. subtilis, B.cereus,* Anthrax vaccination) into their respective pathogen response groups. The identification of the most differentially regulated genes within each pathogen group, facilitated screening for candidate early molecular markers of infection using proteomics analyses. Phase I evaluated three specific secreted cytokines (Amphiregulin, TNF-α and IL-4) and other yet unidentified protein markers that were differentially expressed in specific infections.

g) Future Directions:

Phase II will validate Phase I findings in a larger group of infections. In addition to *E. coli* and *B. cereus,* other common infections, such as those caused by Gram-positive bacteria (*Staphylococcus aureaus,Staphylococcus epidermidis* [coagulase negative], *Streptococcus pyogenes* [Group A, beta hemolytic Strep], *Enterococcus faecalis*) and Gram-negative bacteria (*Pseudomonas aeroginosa, Proteus mirabilis),* virus (*Hepatitis* B) and fungus (*Candida albicans*), will be evaluated. *In vivo* and *in vitro* genetic responses will be correlated and validated and a larger *in vivo/in vitro* database of human response to infections will be generated. Based on gene expression data, sets of protein markers will be identified for specific infections by proteomic analyses. Known and unidentified protein markers will be isolated, identified and characterized for potential coupling to biosensors arrays for rapid detection of exposures to infectious agents in serum or whole blood samples.

Phase II of this study will lead to development of differential biomolecular nano-sensor array systems that measure specific marker proteins and allow almost immediate detection and identification of early differential immune response to specific microbial pathogens. A proposal (Bio-Molecular Nano-Devices/Systems [MOLDICE] for Detecting Early Molecular Markers of Injury, Toxin Exposure and Infection) has been submitted to DARPA (BAA01-42) and is being presented to the Director for final decision on funding. The DARPA proposal is a joint proposal with the Polymer Science and Engineering Branch and the Image and Signal Processing Branch, Naval Air Warfare Center Weapons Division (NAWCWD) at China Lake (NAWCWD is also assisting with Phase II of this project). An electronically addressable array of ion-channel biosensors will be developed for rapid analysis of blood for injury, toxin exposure and infection. This project will initially demonstrate an ion-channel sensor based on α-hemolysin pores and short peptides that mimic physiologic receptors incorporated into stabilized bilayer-lipid membranes. Binding kinetics will identify unique signatures for ligands. This sensor will provide selectivity in complex biological fluids, reversibility of ligand/receptor interaction and measurable changes in ion flux across the pore. Once proof-of-concept is completed, coupling of mimic physiologic receptor peptides to more stable polymer membranes, large-scale integration and parallel array processing of stochastic signals from individual sensing elements will be accomplished.

6) List of Publications and Technical Reports:

a) Publications/Manuscripts submitted: None

b) Technical Reports submitted to ARO:

(1) Guck Ooi, Ph.D.; Sun H. Paik, Ph.D.; Earl W. Ferguson, M.D., Ph.D., Molecular Signature of Biological Pathogens, 7/2002, 8/2002, 9/2002, 10/2002, 11/2002

7) Participating scientific personnel of this project:

Guck Ooi, Ph.D., Monash University, Melbourne, Victoria, 1988; Master of Applied Science, RMIT, Victoria, 1984; Bachelor of Applied Science (Distinction) in Applied Biology, Royal Melbourne Institute of Technology (RMIT), Victoria, 1980.

Sun Paik, Ph.D. in Cell and Molecular Biology, University of Maryland, College Park, 1997 M.S. in Pharmacology, 1991; and D.Pharm., 1987, University of Sao Paulo, Sao Paulo, Brazil.

Earl Ferguson, M.D./Ph.D. (Physiology), University of Texas Medical Branch, Galveston, Texas, 1970; B.A. (Chemistry), Baylor University, Waco, Texas, 1965; Medicine Resident, Cardiology Fellow and Research Associate (Biochemistry), Duke University Medical Center, Durham, North Carolina, 1971-75.

Yoko Murata, Ph.D. in Nuclear Technology, Nuclear Energy Research Institute of São Paulo (IPEN-CNEN/SP), University of São Paulo, 1992-95; M.S. (Nuclear Technology [Radiobiology]), IPEN-CNEN/SP, University of São Paulo, 1984-87; Degree in Pharmacy and Biochemistry, Specialization in Toxicological and Clinical Analysis, University of São Paulo, Brazil, 1979-84.

Steve Horvath, Ph.D. in Mathematics, University of North Carolina, Chapel Hill, 1995; **Sc.D.** in Biostatistics, Harvard School of Public Health, 2000

8) Report of inventions: None

9) Bibliography:

1. Beutler B. Sepsis begins at the interface of pathogen and host. Biochem. Soc. Trans. 29: 853-9, 2001.
2. Manger ID, and Relman DA. How the host sees pathogens: global gene expression responses to infection. Curr. Opinion Immunol. 12: 215-8, 2000.
3. Kellam P. Host-pathogen studies in the post-genomic era. Genome Biol. 1(2): reviews 1009.1-1009.4, 2000.
4. Diterich I, Harter L, Hassler D *et al.* Modulation of cytokine release in ex vivo-stimulated blood from borreliosis patients. Infect. Immun. 69: 687-94, 2001.
5. L. Breiman. Prediction games and arcing algorithms. Neural Comput. 11(7): 1493-517, 1999.

10) Appendix:

See Attached Figures.

Figure 1. Cluster of all samples with top 1000 most varying genes:

All genes/42 samples

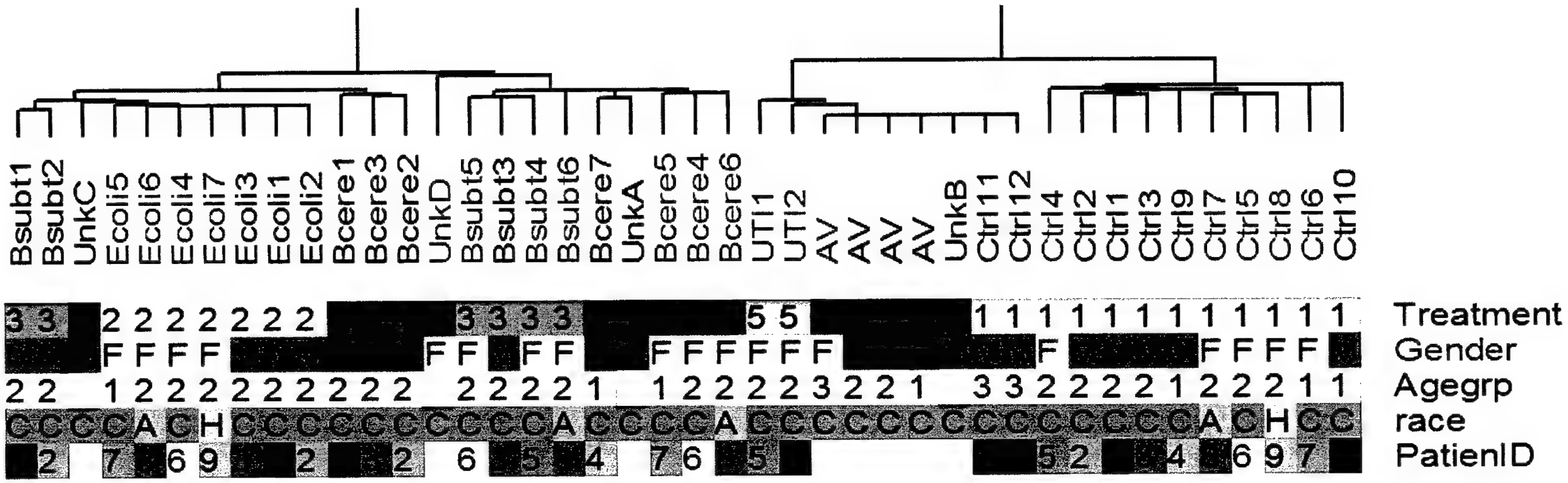


Figure 2. Unsupervised Analysis
Multidimensional Scaling Plots (MDS plots)

MDS for all samples with 1000 most varying genes

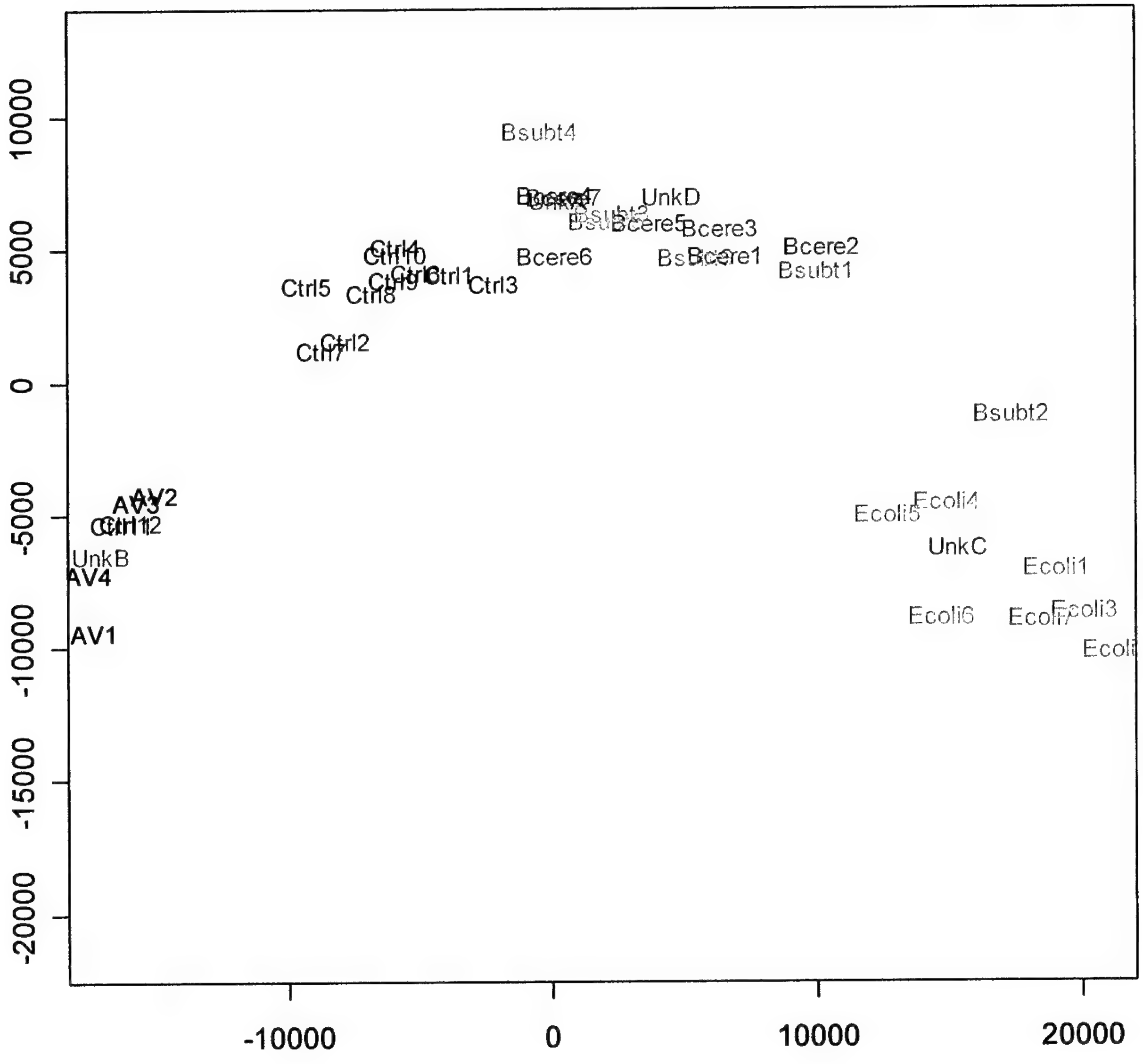


Figure 3A. Top 20 genes that best separate Control, E. coli, B. subtilis and B. cereus

			Basic Statistics for B.cereus					Basic Statistics for B.subtilis					Basic Statistics for Control					Basic Statistics for E.coli			
probe set	Gene	N	MIN	MAX	MEAN	STD	N	MIN	MAX	MEAN	STD	N	MIN	MAX	MEAN	STD	N	MIN	MAX	MEAN	STD
201422_at	interferon, gamma-inducible protein 30	7	2247.19	4252.29	3018.519	777.6964	6	437.852	3486.87	1894.414	1358.885	12	4653.21	6697.61	5591.279	755.8779	7	807.268	2178.08	1519.315	562.9388
202313_at	protein phosphatase 2 (formerly 2A), regulatory subunit I	7	750.531	915.199	814.4891	56.28703	6	655.446	771.931	703.5432	48.14305	12	569.308	799.133	670.828	67.80633	7	913.062	1071.44	998.245	56.890191
202423_at	zinc finger protein 220	7	1710.99	2251.9	1929.284	178.9759	6	2119.7	2683.82	2518.537	206.8486	12	1314.82	2487.39	1969.293	357.3924	7	1594.04	2213.31	1840.491	237.45234
203725_at	growth arrest and DNA-damage-inducible, alpha	7	1317.25	1544.15	1440.093	97.42755	6	861.871	1288.09	1022.539	142.3813	12	106.204	397.403	244.8137	84.67264	7	550.489	910.182	656.2057	134.44968
204094_s_at	KIAA0669 gene product	7	658.401	1059.62	938.0267	136.9585	6	301.147	698.949	536.9717	154.4828	12	165.502	514.744	380.6585	115.7967	7	273.94	606.324	458.4423	109.93329
204747_at	interferon-induced protein with tetratricopeptide repeats	7	115.969	896.495	357.7609	263.0266	6	235.634	2896.09	1063.198	928.703	12	62.7649	282.305	157.7664	64.0446	7	3337.76	5385.35	4633.243	698.12239
205266_at	leukemia inhibitory factor (cholinergic differentiation factc	7	190.706	413.519	315.6036	79.24338	6	1.88007	209.458	89.75028	72.23908	12	-12.0565	71.5669	16.2205	23.27183	7	39.7272	219.299	123.6592	60.826383
206134_at	ADAM-like, decysin 1	7	45.5023	70.5485	57.38736	9.138366	6	41.8332	68.8517	52.44625	11.44002	12	35.8212	77.9829	56.721	12.48965	7	141.8	250.944	186.095	38.702649
206181_at	signaling lymphocytic activation molecule	7	308.24	920.818	551.2363	256.6667	6	964.121	1602.87	1233.432	231.9263	12	92.9451	283.734	155.5611	55.12341	7	348.766	1557.09	799.0627	409.56777
207067_s_at	histidine decarboxylase	7	34.8674	72.1625	55.9411	14.66443	6	31.5238	80.8737	59.93982	16.95542	12	145.175	396.807	221.3271	83.70721	7	214.349	512.222	306.6254	101.21025
207270_x_at	CMRF35 leukocyte immunoglobulin-like receptor	7	84.9731	148.321	117.2628	23.68121	6	28.2341	146.519	89.3305	48.68436	12	182.934	414.658	269.4833	76.8418	7	61.1525	106.468	79.99967	14.573648
207375_s_at	interleukin 15 receptor, alpha	7	59.0208	141.42	97.69974	33.07193	6	113.256	279.892	162.9	60.39423	12	36.5521	118.076	58.68496	23.50954	7	375.249	623.86	511.1433	90.968216
211751_at	similar to rat myomegalin	7	42.13	71.7322	50.9901	10.29525	6	13.5205	31.991	22.64537	8.688046	12	24.5629	68.8077	41.28492	10.91492	7	32.1227	81.7092	57.52217	21.462371
212655_at	KIAA0579 protein	7	180.203	309.32	238.746	50.38756	6	344.42	510.176	444.7697	66.37042	12	111.756	442.336	266.881	96.55148	7	125.857	363.725	208.8407	95.810925
214933_at	calcium channel, voltage-dependent, P/Q type, alpha 1A	7	70.8206	88.9192	77.40869	6.532646	6	83.2997	111.98	95.32823	10.64628	12	53.7191	84.8897	72.82529	8.878301	7	251.762	381.894	296.6827	43.527619
216020_at	Consensus includes gb:AL080107.1 /DEF=Homo sapien	7	25.4348	60.9827	36.06693	12.01393	6	16.141	74.008	38.7349	19.16951	12	1.43762	22.4432	10.16487	6.687167	7	70.0699	188.473	135.4941	46.611124
217741_s_at	zinc finger protein 216	7	1710.98	2478.53	1991.243	244.9371	6	1079.01	1386.12	1219.917	127.4788	12	793.782	1524.84	1174.009	240.3398	7	976.084	1316.15	1133.839	138.76702
219159_s_at	19A24 protein	7	278.628	688.834	515.5854	148.021	6	335.432	690.165	517.668	115.5151	12	276.986	822.083	469.0824	145.957	7	900.87	1577.25	1149.564	249.58536
220306_at	hypothetical protein FLJ20202	7	973.871	1364.3	1217.094	135.4559	6	682.212	1229.74	886.2957	190.2464	12	153.737	509.981	336.6083	96.61883	7	728.48	956.205	841.7569	100.04928
51146_at	hypothetical protein FLJ20477	7	110.434	237.386	183.5219	42.97177	6	184.629	282.748	249.9078	34.21281	12	110.856	240.212	165.7603	43.96001	7	117.104	211.087	167.9116	36.057563

Figure 3B. Cluster of all samples with top 20 genes obtained Random Forest Prediction and step-wise linear discriminant analysis

20 genes/34 samples

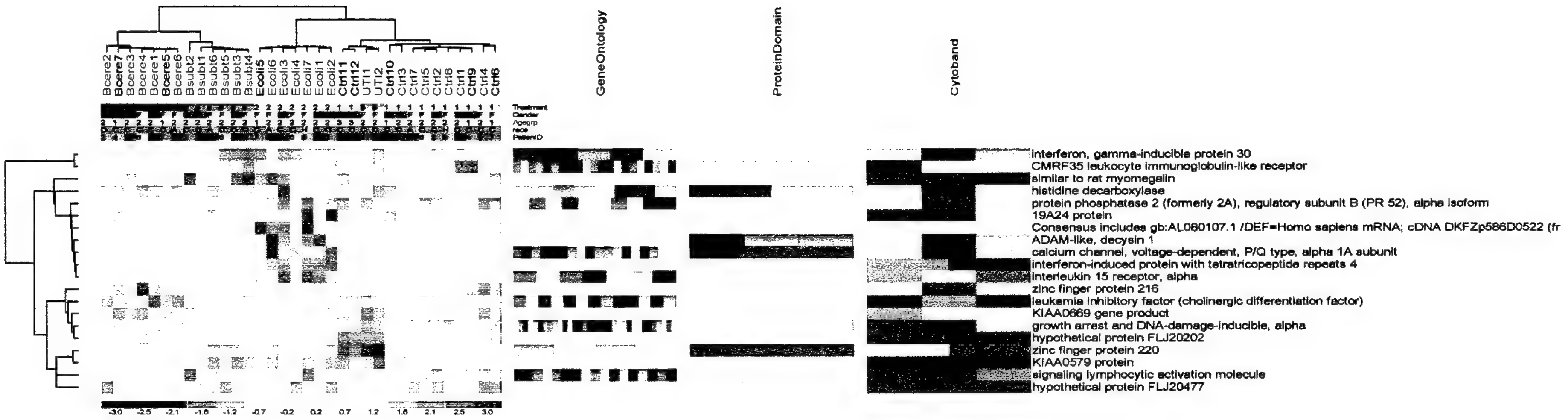


Figure 3C. Box-plots of the expression of top 20 genes separating TRT

To understand where these 20 most important genes are over expressed, we show here the box-plots versus pathogen status of the most important genes.

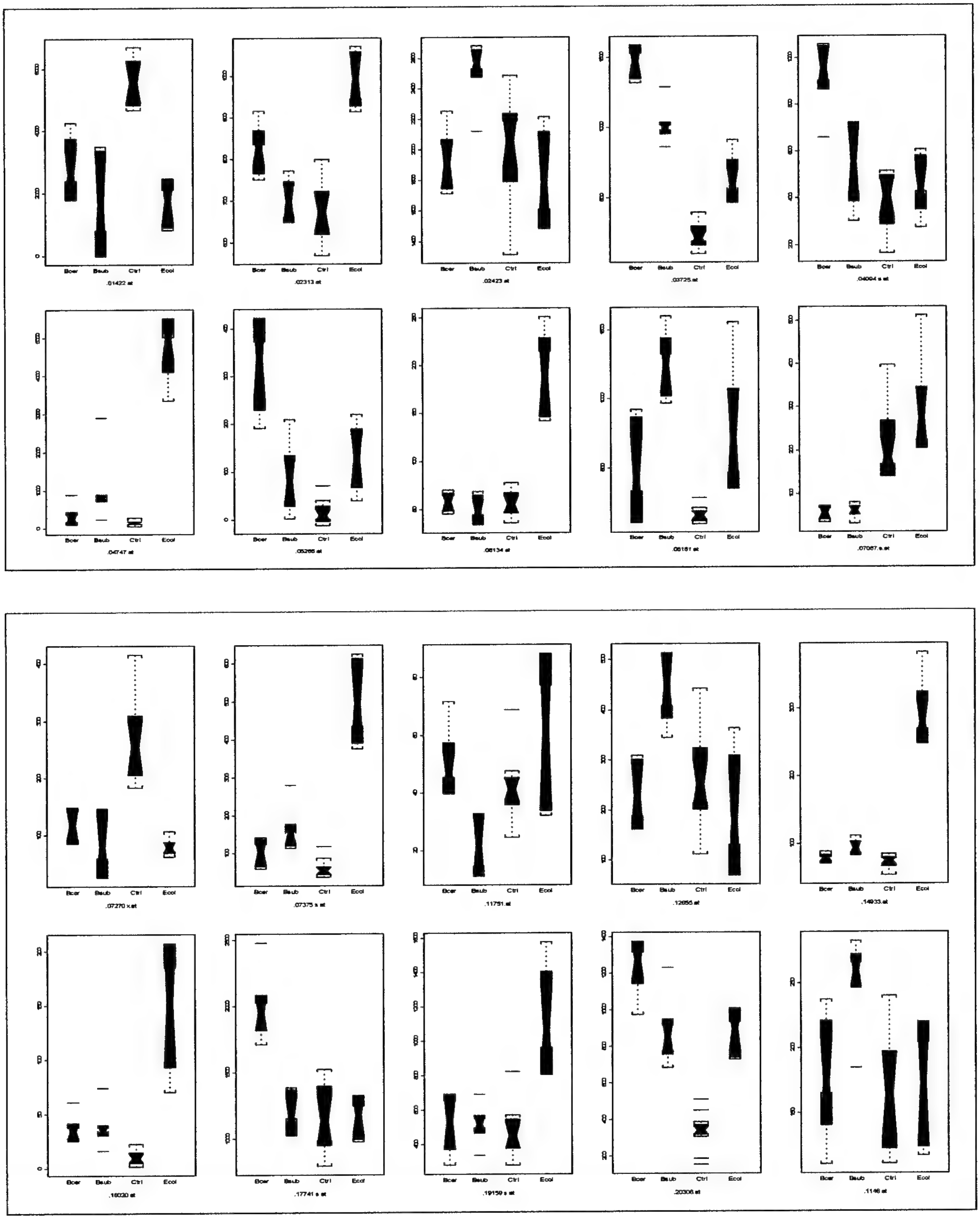
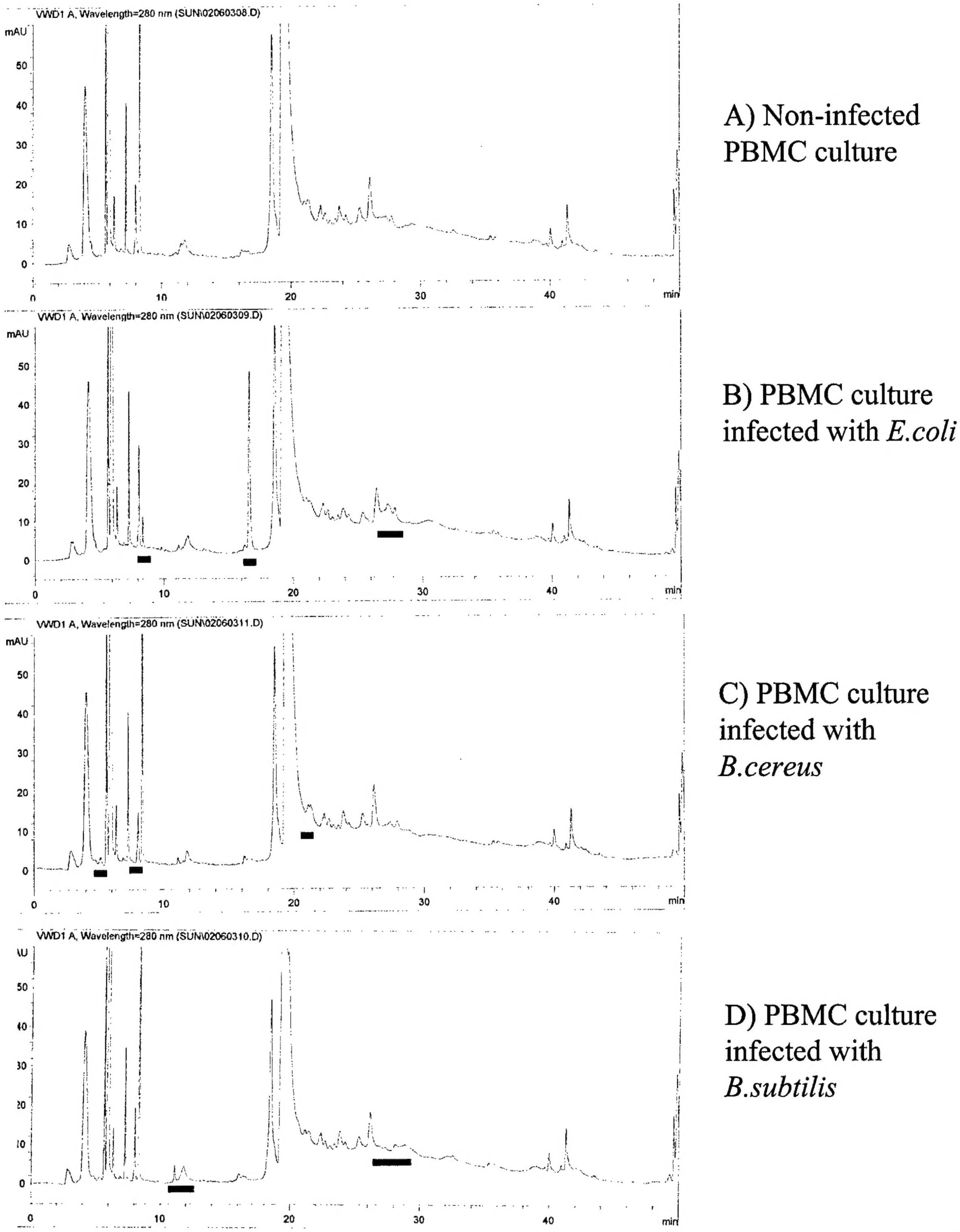


Figure 4A. HPLC profiles of PBMC culture supernatants of non-infected and 3 h after infection with *E. coli, B. cereus,* and *B. subtilis*. Differentially secreted proteins by specific infections are underlined. Sample load: 100 µl of culture supernatants containing 10% serum. Chromatography conditions: Linear gradient from 20 to 70% acetonitrile containing 0.1% TFA over 30 min with Vydac C18 column, flow rate of 1.2 ml/min, with detection at 280 nm. The profiles A to D represent same subject samples before and after infection and protein peaks underlined are representative of 3-5 samples.



A) Non-infected PBMC culture

B) PBMC culture infected with *E.coli*

C) PBMC culture infected with *B.cereus*

D) PBMC culture infected with *B.subtilis*

Figure 4B. HPLC profile of the same subject plasma before (A) and 24 h after Anthrax vaccination (B). Same chromatographic conditions described in figure A were used. Differentially secreted proteins are underlined.

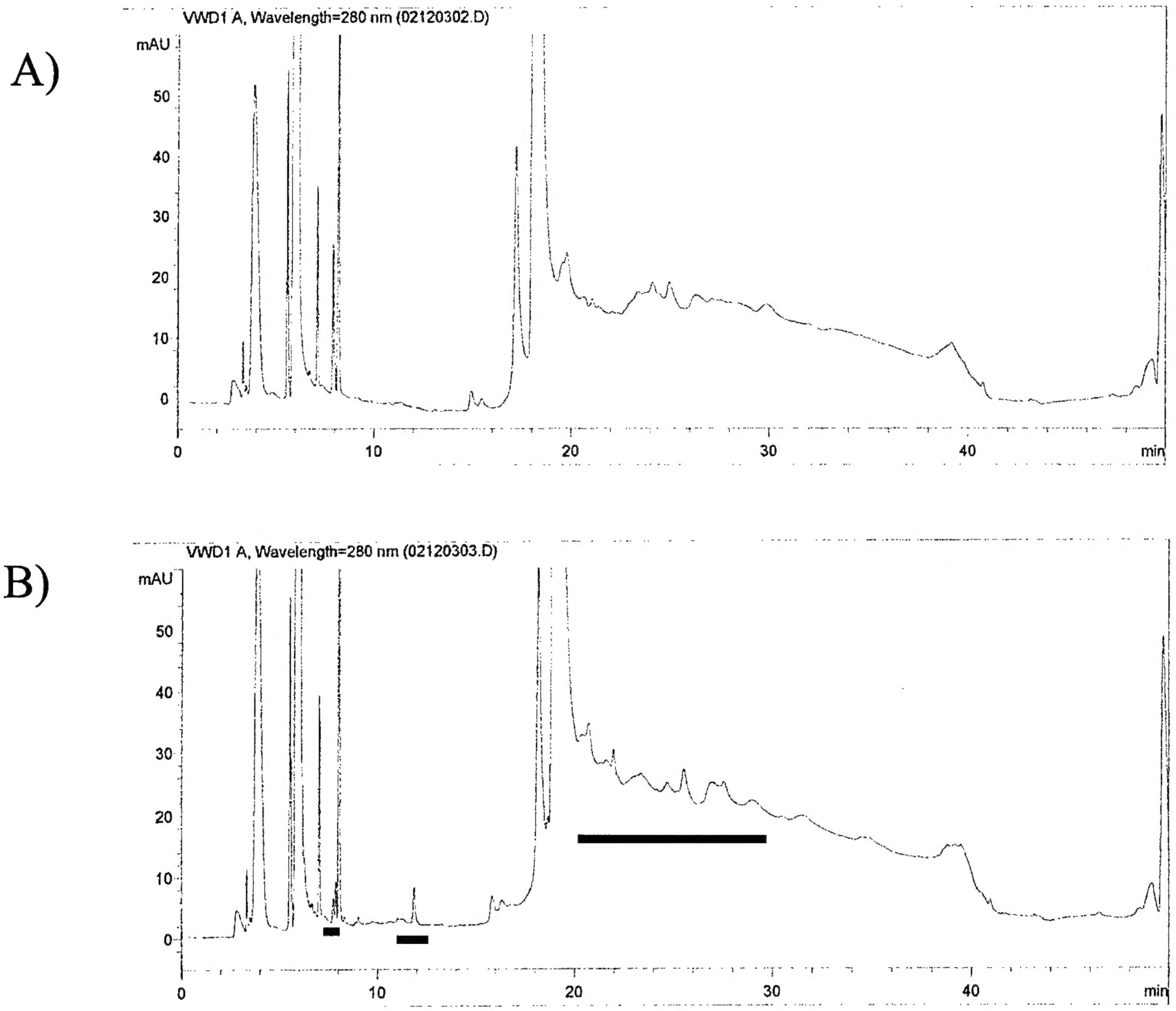


Figure 5. 2D electrophoresis gel analysis of PBMC culture supernatants before (A) and 3 h after infection with *E.coli* (B) and *B.cereus* (C). Gel A shows basal proteins secreted in the absence of infection. First dimension separation was by IEF from pH 4-7 in an IPG gel. Second dimension separation was by SDS-PAGE in an 8-16% T Polyacrylamide gradient gel. Gels were stained with SyproRuby stain. Some of differentially expressed protein spots are circled.

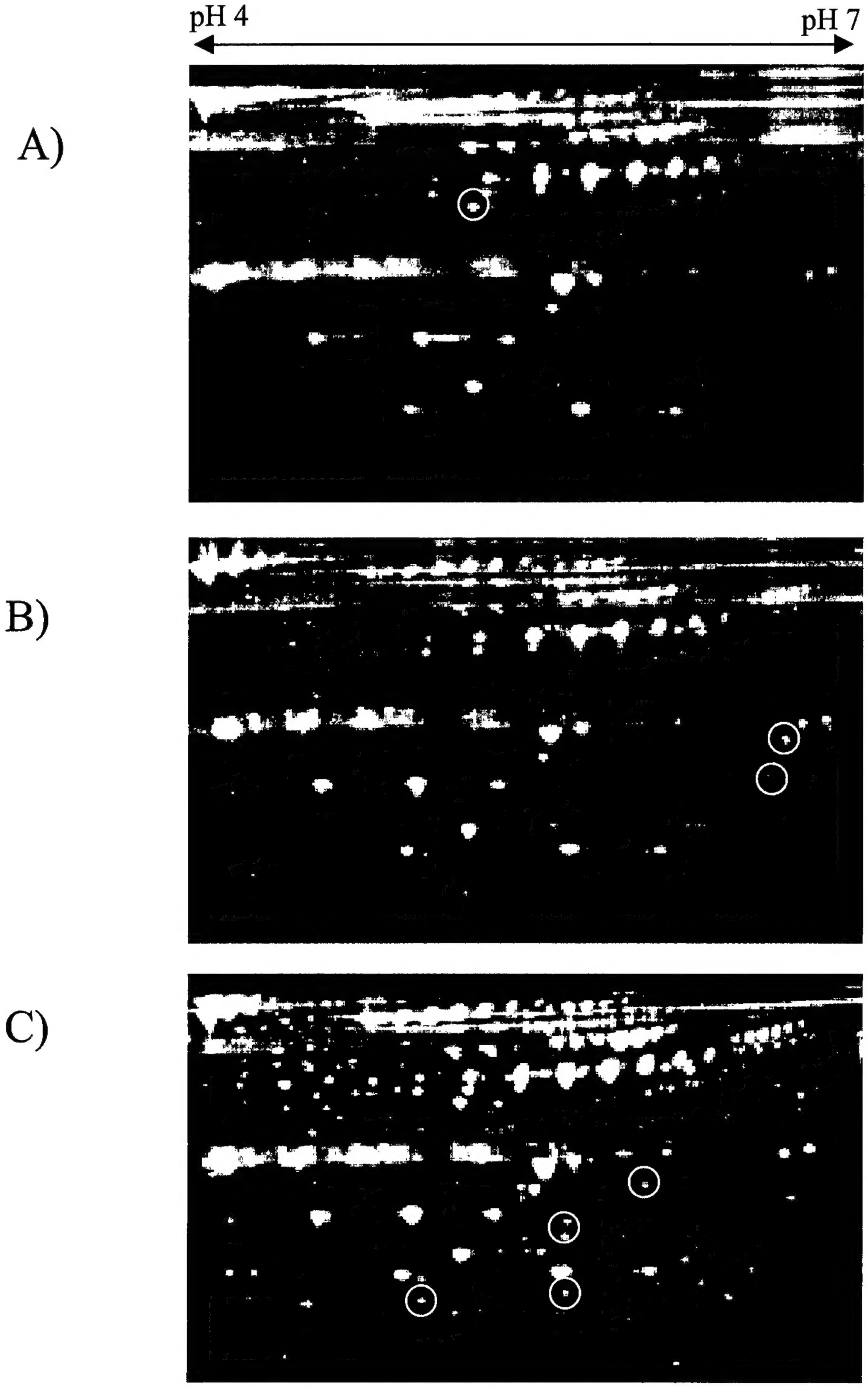


Figure 6. Western blot analysis of PBMC culture supernatants of non-infected control group (C) and 3 h after *in vitro* infection by *E.coli* (EC), *B.subtilis* (BS), *B. cereus* (BC). Levels of TNF-α and IL-4 secreted 3 h after infection correlate with DNA microarray gene expression data. Only trace amounts of amphiregulin were detected on subjects 2 and 3 under *E.coli* infection, not under *Bacillus* sp. infections according to gene expression data.